Theor Appl Genet (2016) 129:2003–2017
DOI 10.1007/s00122-016-2755-6

ORIGINAL ARTICLE

Fine mapping of *Restorer-of-fertility* in pepper (*Capsicum annuum* L.) identified a candidate gene encoding a pentatricopeptide repeat (PPR)-containing protein

Yeong Deuk Jo[1,2] · Yeaseong Ha[1] · Joung-Ho Lee[1] · Minkyu Park[1] · Alex C. Bergsma[3] · Hong-Il Choi[2] · Sandra Goritschnig[3] · Bjorn Kloosterman[3] · Peter J. van Dijk[3] · Doil Choi[1] · Byoung-Cheorl Kang[1]

Received: 1 April 2016 / Accepted: 15 July 2016 / Published online: 28 July 2016

Abstract

Key message **Using fine mapping techniques, the genomic region co-segregating with *Restorer-of-fertility* (*Rf*) in pepper was delimited to a region of 821 kb in length. A PPR gene in this region, *CaPPR6*, was identified as a strong candidate for *Rf* based on expression pattern and characteristics of encoding sequence.**

Abstract Cytoplasmic-genic male sterility (CGMS) has been used for the efficient production of hybrid seeds in peppers (*Capsicum annuum* L.). Although the mitochondrial candidate genes that might be responsible for cytoplasmic male sterility (CMS) have been identified, the nuclear *Restorer-of-fertility* (*Rf*) gene has not been isolated. To identify the genomic region co-segregating with *Rf* in pepper, we performed fine mapping using an *Rf*-segregating population consisting of 1068 F_2 individuals, based on BSA-AFLP and a comparative mapping approach. Through six cycles of chromosome walking, the co-segregating region harboring the *Rf* locus was delimited to be within 821 kb of sequence. Prediction of expressed genes in this region based on transcription analysis revealed four candidate genes. Among these, *CaPPR6* encodes a pentatricopeptide repeat (PPR) protein with PPR motifs that are repeated 14 times. Characterization of the CaPPR6 protein sequence, based on alignment with other homologs, showed that *CaPPR6* is a typical *Rf*-like (*RFL*) gene reported to have undergone diversifying selection during evolution. A marker developed from a sequence near *CaPPR6* showed a higher prediction rate of the *Rf* phenotype than those of previously developed markers when applied to a panel of breeding lines of diverse origin. These results suggest that *CaPPR6* is a strong candidate for the *Rf* gene in pepper.

Introduction

Cytoplasmic-genic male sterility (CGMS) has been used in commercial hybrid seed (F_1) production in a large number of crops (Hanson and Bentolila 2004). In hybrid seed production using CGMS, a male-sterile line carrying a maternally inherited cytoplasmic male sterility (CMS)-inducing gene located on the mitochondrial genome is used as a seed parent line. This enables prevention of self-pollination without laborious emasculation during seed production. Meanwhile, a line carrying a *Restorer-of-fertility* (*Rf*) gene, nucleus-encoded gene that suppress the CMS-inducing gene, is used as an pollen parent. The resulting F_1 plant is male-fertile due to the action of *Rf* although the F_1 plant has the CMS-inducing gene.

Efforts to clone and characterize *Rf* genes have been made in many crop species because of their economic

Communicated by M. J. Havey.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-016-2755-6) contains supplementary material, which is available to authorized users.

✉ Byoung-Cheorl Kang
bk54@snu.ac.kr

[1] Department of Plant Science, Plant Genomics and Breeding Institute, and Research Institute for Agriculture and Life Sciences, Seoul National University, 599 Gwanak-ro, Gwank-gu, Seoul 151-921, Korea

[2] Advanced Radiation Technology Institute, Korea Atomic Energy Research Institute, Jeongeup 580-185, Republic of Korea

[3] Keygene N.V., Wageningen 216, The Netherlands

value in efficient hybrid seed production, as well as their scientific importance as a model to analyze nuclear control of mitochondrial gene expression, and co-evolution of the nuclear genome with mitochondrial DNA (Chen and Liu 2014; Hanson and Bentolila 2004; Fujii et al. 2011; Kubo et al. 2011). As a result, *Rf* genes have been successfully cloned in maize (Cui et al. 1996), petunia (Bentolila et al. 2002), radish (Brown et al. 2003; Desloire et al. 2003; Koizuka et al. 2003), rice (Hu et al. 2012; Huang et al. 2015; Kazama and Toriyama 2003; Komori et al. 2004; Luo et al. 2013; Wang et al. 2006), and sugar beet (Matsuhira et al. 2012).

The majority of cloned *Rf* genes are members of the pentatricopeptide repeat (PPR) gene family. Through genetic mapping, co-segregation of PPR genes with *Rf* has also been identified in several crops including sorghum, *Mimulus*, and maize (CMS-S) (Barr and Fishman 2010; Klein et al. 2005; Xu et al. 2009). However, four non-PPR *Rf* genes have also been cloned, encoding an aldehyde dehydrogenase (Rf2a), a glycine-rich protein (Rf17), a putative retrograde signaling control-related protein (Rf2), and a putative mitochondrial protein quality control-related protein (Rf1), in CMS-T maize, CW-CMS rice, LD-CMS rice, and sugar beet, respectively (Cui et al. 1996; Fujii and Toriyama 2009; Itabashi et al. 2011; Matsuhira et al. 2012).

PPR genes encode proteins with repeated motifs consisting of degenerate arrays of 35 amino acids of slightly different lengths. They can bind RNA through their superhelix structure (Small and Peeters 2000). Depending on their structure, which includes P and PLS-class subfamilies, PPR genes have been shown to be involved in RNA editing, splicing, processing and degradation (Barkan and Small 2014; Lurin et al. 2004). Most PPR-type *Rf* genes are in the P subfamily, which is characterized by an array of regular 35-amino acid motifs (Bentolila et al. 2002; Brown et al. 2003; Komori et al. 2004). The products of *Rf* genes of this type are involved in the processing or degradation of transcripts from CMS-expressed genes, as exemplified by Rf1a and Rf1b, which function in cleavage or degradation of CMS-associated transcripts (*atp6-orf79*) in rice (Wang et al. 2006).

PPR genes constitute a large gene family (e.g., 441 in Arabidopsis and 477 in rice). They are present only in land plants, although a few homologs have been detected in other species such as yeasts and algae (Fujii et al. 2011). The cloned PPR-type *Rf* genes typically form a cluster with closely located PPR genes, whereas other PPR genes are dispersed throughout the genome (Lurin et al. 2004; Wang et al. 2006). The *Rf* and clustered PPR genes share high sequence similarity; based on phylogenetic analysis using known PPR genes from a broad range of plant species, Fujii et al. (2011) classified these genes as *Rf*-like (*RFL*) genes. *RFL* genes from diverse plant species form a separate clade from other PPR genes. This implies that *RFL* genes originated from the same ancestral gene that existed before the speciation of land plants. Finally, *RFL* genes show much higher rates of non-synonymous to synonymous substitutions than other PPR genes (Fujii et al. 2011). The rate of diversifying selection is highest on the first, third and sixth amino acid of the PPR motif, which may be involved with RNA ligand interaction. Studies of PPR genes have shown that each of their 35 amino acid arrays determines the specificity of the protein to one nucleotide of target RNA. Furthermore, a combination of the first and sixth amino acids in the motif is crucial for determining specificity, implying that these amino acids interact directly with RNA (Barkan et al. 2012; Barkan and Small 2014). These characteristics of *Rf* genes support the hypothesis that they have evolved by a birth-and-death process usually found in disease resistance genes, to cope with the evolution of mitochondrial *orf*s including CMS genes (Dahan and Mireau 2013; Touzet and Budar 2004).

In pepper, only a CMS cytoplasm originated from an Indian accession (USDA accession PI 164835) has been characterized (Peterson 1958) and used in hybrid seed production. A chimeric gene (*orf507*) and a pseudo gene (*Ψatp6-2*) have been isolated and characterized as strong candidates to be CMS-associated genes (Gulyas et al. 2006; Kim and Kim 2006; Kim et al. 2007; Li et al. 2013). However, for *Rf*, only molecular markers have been developed, based primarily on random amplified polymorphic DNA (RAPD) or amplified fragment length polymorphism (AFLP) methods using random primers (Kim 2005; Kim et al. 2006; Lee et al. 2008a; Min et al. 2009). Additional markers linked to *Rf* were developed from end sequences of bacterial artificial chromosome (BAC) clones containing sequences homologous to petunia *Rf*, or by comparative mapping using tomato *RFL* genes. This implies that the pepper *Rf* gene may be an *RFL*-type PPR gene (Jo et al. 2010). A 2013 study showed the differential expression of several PPR genes between the restorer and its near-isogenic CMS line in pepper (Liu et al. 2013). Although several markers were closely linked to the *Rf* gene (e.g. OPP13: 0.4 cM, AFRF4: 0.1 cM; Min et al. 2009) in a given population, most of the markers were unable to correctly detect genotypes for a wider range of pepper lines, implying that rearrangements have occurred in the *Rf* locus (Jo et al. 2010). To overcome this limitation, the cloning of the *Rf* gene itself and development of markers based on co-segregating sequences are required.

In this study, we identified a genomic region co-segregating with the restoration-of-fertility phenotype, and cloned a candidate for the *Rf* gene through the use of combinational mapping strategies. These results should lead to more reliable and efficient molecular breeding for *Rf* and provide greater understanding of the interaction between *Rf* and CMS cytoplasm in pepper.

Materials and methods

Plant materials, phenotyping, and test crosses

Two different F_2 populations were used for molecular marker linkage analysis. A total of 1068 F_2 plants—derived from a previously developed cross between a CMS line (S/*rfrf*) and a restorer line (N/*rfrf*) which were seed and pollen parents of a 'Chungyang' cultivar and named as 'Chungyang A' and 'Chungyang C' in this study, respectively (Monsanto Korea, Chungju, Korea) (Jo et al. 2010)—were used for fine mapping of *Rf* (Table S1). Another F_2 population, described by Kim et al. (2013), was composed of 160 F_2 plants and derived from a cross between a CMS line 'TCMS' (S/*rfrf*), and a restorer line 'Milyang K' (S/*RfRf*) (Table S1). Leaf samples from a total of 51 breeding lines provided by Monsanto Korea (Jo et al. 2010), and 50 lines from Enza Zaden (Enkhuizen, The Netherlands), were used to validate the developed markers. For the transcriptome analysis, anthers from a 'Bukang' seed parent 'Bukang A' (S/*rfrf*) and a pollen parent named as 'Bukang C' (N or S/*RfRf*), were used.

The fertility of F_2 individuals was evaluated at least three times by careful inspection for the presence of pollen, as described previously (Jo et al. 2010; Kim et al. 2013). Monsanto Korea (Jo et al. 2010) and Enza Zaden provided phenotypes for the breeding lines.

To determine the *Rf* genotype of Criollo de Morelos 334 (CM334) from which the pepper scaffold and BAC sequences used in this study originated, crosses CM334 X Milyang A (a CMS line) and CM334 X Bukang A (a CMS line) were performed. Fertility was investigated in five F_1 plants from each cross.

Amplified fragment length polymorphism (AFLP)

Tentative *Rf* genotypes of individuals from a TCMS × Milyang K F_2 population (160 individuals) were determined based on plant fertility and the genotypes of the *Rf*-linked markers G05G1-HRM and OPP13C1-CAPS (Jo et al. 2010; Kim et al. 2013). Two different pairs of DNA pools consisting of DNA from ten individuals with *RfRf* and *rfrf* genotypes, respectively, were prepared. Using these DNA pools, bulked segregant analysis (BSA)-AFLP was carried out for 1024 primer combinations covering all combinations between *Eco*RI + ANN or CNN, and *Mse*I + ANN or CNN. Primer combinations with clear polymorphisms were selected. To reduce efforts to apply all the selected primer combinations to all F_2 individuals, further selection of primer combinations was performed by applying them to DNA pools of recombinants which were screened by application of only eight primer combinations randomly chosen among the selected primer combinations to all F_2 individuals (160 individuals). The primer combinations in which marker genotype was consistent with fertility restoration phenotype of individuals in each recombinant pool were applied to all F_2 individuals to confirm co-segregation with the phenotype. AFLP analysis was carried out according to a protocol developed by KeyGene (Wageningen, The Netherlands), modified from the methodology reported by Vos et al. (1995). Re-amplification of amplicons from selected AFLP markers and Sanger sequencing of PCR products, were conducted to determine marker sequences.

Marker development based on tomato gene sequences

Thirteen expressed sequence tags (ESTs) from pepper, which showed high similarity to tomato genes located in the 1.3–2.2 Mb region of tomato chromosome 6, were selected from an EST database of *C. annuum* 'Bukang' (Kim et al. 2008; Tomato Genome C 2012). Putative positions of introns for pepper ESTs were predicted by comparing them with the tomato gene DNA sequences. Primers were designed from EST sequences to amplify predicted intron sequences. PCR using the designed primers was performed for Chungyang A and Chungyang C. Those primer combinations generating amplicons in only one of the two lines were used as sequence characterized amplified region (SCAR) markers. When amplification occurred in both lines, amplicon sequences were compared to develop cleaved amplified polymorphic sequence (CAPS) markers or markers based on high resolution melting (HRM) analysis.

Application of DNA markers

For SCAR markers, PCR reactions were carried out in 20-μl reactions with 50 ng template DNA, 200 μM deoxynucleotide solution mix (dNTP), 2 μl 10× *Taq* DNA polymerase buffer, 5 pmol each primer, and 1 unit DNA *Taq* DNA polymerase. PCR amplification was performed using an initial cycle of 94 °C for 5 min; 35 cycles of 94 °C for 30 s, annealing temperature of each marker (Table S2) for 30 s and 72 °C for 2 min; followed by a final extension of 72 °C for 10 min. PCR products (5 μl each) were separated on 0.8 % agarose gels. For CAPS markers, the PCR amplification procedure was the same as that for SCAR markers, except it was followed by digestion of PCR products with 5 units of suitable restriction enzyme (Table S2) at 37 °C for 3 h. Digestion products were separated on 1.5 % agarose gels. For HRM analysis, the procedure described by Jeong et al. (2010) was applied.

Screening pepper scaffold sequences containing sequences of developed markers

Scaffold sequences of the *C. annuum* CM334 genome were kindly provided by the Horticultural Crop Genomics laboratory (Seoul National University, Seoul, Republic of Korea) in an earlier stage of fine mapping. Scaffolds containing marker sequences were selected by BLAST searches using the Pepper Genome Platform website (http://peppergenome.snu.ac.kr). Version 0.83 contigs and version 0.9 scaffolds were used to develop and anchor markers, respectively. Later, scaffold sequences included in the DNA region co-segregating with *Rf* were replaced with up-to-date versions of these chromosome sequences (version 1.55) (Kim et al. 2014).

BAC library screening, chromosome walking, and sequencing selected BAC clones

A CM334 BAC library developed by Yoo et al. (2003) was strategically pooled for efficient screening of BAC clones based on PCR analysis. A three-step BAC screening process was developed to include pooling of cell cultures from each column on each 384-well plate, pooling of cell cultures on the same 384-well plate, and two-dimensional pooling of the cell culture pooled from each plate. This BAC library pooling strategy enabled a single copy of a target BAC clone to be selected from among 221,184 clones using three rounds of PCR (140 PCR reactions).

Six chromosome walking cycles were carried out, each cycle including: (1) design of a primer set to amplify a specific sequence located on the distal end of former sequences; (2) screening new BAC clones with a PCR-based BAC library screening system using the designed primer set; (3) BAC end sequencing for the selected BAC clones; and (4) elucidation of the overlap pattern between selected BAC clones, based on PCR amplification of BAC end sequences using BAC clone DNA as a templates. BAC clones with the longest extension from the former tiled sequences were selected for the next cycle of tiling.

The GS-FLX system (Roche Applied Science, Indianapolis, USA) was used to sequence the selected BAC clones. Each BAC clone was given a different tag during library construction for separate assembly. Generated sequences were assembled using Newbler Assembler Software Version 2.0 (454 Life Sciences, Branford, USA) at the National Instrumentation Center for Environmental Management (NICEM; Seoul, Republic of Korea). Gaps between BAC clone sequence contigs were filled with Sanger-sequenced PCR amplicons of gap sequences using primer sets designed from the distal ends of contigs and on the released pepper chromosome sequences (version 1.55) (Kim et al. 2014).

Transcriptome analysis and analysis of candidate genes

RNA was isolated from anther tissues of restorer line Bukang C and CMS line Bukang A using the Hybrid-RTM RNA extraction kit (GeneAll Biotechnology, Seoul, Republic of Korea) according to the manufacturer's instructions. Transcriptome sequences were produced using Hiseq 2500 (Illumina/Solexa, San Diego, USA) at NICEM.

To determine candidate genes, repeated sequences on the genomic region co-segregating with *Rf* were masked, and RNA-Seq reads were mapped on the unmasked euchromatic region. Repeated sequences were visualized using the MultiPipMaker program (available at: http://pipmaker.bx.psu.edu/pipmaker/). Sequence regions showing similarity (*E* value $<10^{-5}$) with transposable element sequences released from the Pepper Genome Database (Qin et al. 2014; available at: http://peppersequence.genomics.cn/) were BLAST-screened and visualized using an in-house Perl script. CLC Workbench 8.5 (CLC bio, Aarhus, Denmark) was used to map NGS reads; read mapping and large-gap mapping functions of this software were used to map reads with greater than 98 % similarity to the genomic sequence co-segregating with *Rf*. De novo assembly and transcript discovery functions in CLC Workbench 8.5 were used to assemble reads mapped by read mapping and large-gap mapping functions, respectively. Assembled contigs were curated manually. Only the contigs on unmasked genomic regions were selected for further analysis.

BLASTX was used to analyze contig sequences, comparing them with the GenBank protein database (http://www.ncbi.nlm.nih.gov/genbank/) and the ITAG *Solanum lycopersicum* protein database (v2.3) (http://solanaceae.plantbiology.msu.edu/). After mapping RNA-Seq reads to coding sequences using CLC Workbench 8.5, heat maps showing the expression levels of candidate genes were drawn using an in-house Perl script.

Cloning of genes homologous with *PPR6*

Forward and reverse primers designed from sequences located 5′ upstream and 3′ downstream of *CaPPR6*, respectively, were used for PCR (forward: 5′-GTGCTGCCTG CTGGTGAAGAA-3′, reverse: 5′-GTATCACAGCGATA GCAGAGCC-3′). PCR amplicons were cloned using the pGEM®-T Easy vector system (Promega, Madison, USA) and randomly selected colonies were sequenced. Translated sequences of *CaPPR6* and homologous genes were aligned using the ClustalW program in CLC Workbench 8.5.

RT-PCR analysis

Using the Hybrid-R™ RNA extraction kit (GeneAll Biotechnology, Seoul, Republic of Korea) according to the

manufacturer's protocol, total RNA was isolated from stems, leaves, ovules, and anthers (obtained from 3 to 5 mm floral buds) from two individuals in the Chungyang F_2 population with *RfRf* and *rfrf* genotypes, respectively. cDNA was synthesized from 2 μg total RNA using the MMLV reverse transcription kit (Promega, WI, USA). For RT-PCR, a primer set was designed (using polymorphisms between the gene sequence of *CaPPR6* and homologous sequences of *CaPPR6* in the CM334 genome which were screened by BLAST analysis via the Pepper Genome Platform website) to amplify a portion of the *PPR6* gene in restorer parents of Chungyang and Bukang, but not in CMS parents of Chungyang and Bukang. The sequences of forward and reverse primers were 5′-GCCTCTTCCTTCTCTTCATGTC-3′ and 5′-GCAGAAGGCATCAATAACACCA-3′, respectively. The nucleotide on the 3′ end of the reverse primer was specifically matched with *CaPPR6* in CM334, and to reduce nonspecific amplification, the third nucleotide from the 3′ end was designed not to match with *CaPPR6* nor its homologs. Reverse-transcriptase PCR (RT-PCR) was performed in 50-μl reactions with 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.5 mM each dNTP, 10 pmol each primer, 1 μl synthesized cDNA, and 1 U rTaq polymerase (Takara, Shiga, Japan).

Results

Development of AFLP markers closely linked to the *Rf* gene

To select for AFLP markers closely linked to the pepper *Rf* gene, three BSA-AFLP cycles were carried out using 1024 primer combinations on an *Rf*-segregating F_2 population (160 individuals) derived from a cross between TCMS (CMS line) and Milyang K (restorer line). Based on the results of BSA-AFLP using two different pairs of DNA pools for *RfRf* and *rfrf* genotypes, 92 primer combinations were selected. In the final BSA-AFLP cycle, recombinants on both sides of *Rf* were screened by applying eight primers randomly selected from the 92 primer combinations on 160 F_2 individuals. AFLP was then carried out on pools of recombinants to screen markers co-segregating with the *Rf* phenotype. After confirming co-segregation in 160 F_2 individuals, a total of 12 AFLP markers were selected (Table S1; Fig. S1a), eight of which showed clear genotyping results for all F_2 individuals. These AFLP markers were regarded as markers co-segregating with *Rf* in this population. Although no conflicts were detected between marker genotypes and dominantly scored phenotypes, recombination events between markers were found as, in most cases, marker genotypes were co-dominantly scored. Eight markers were classified into four groups, according to the recombination patterns between them. One of the developed markers, AF5, was identical to AFRF4, previously reported to be located 0.1 cM from *Rf* (Min et al. 2009; Fig. S1b).

Development of markers by comparative mapping using the tomato whole genome sequence

Genetic marker G05G1-HRM was previously developed from a tomato sequence (Jo et al. 2010); based on synteny studies between tomato and pepper at the G05G1-HRM genome region, it was possible to design additional markers. Three markers (1.31-SCAR, 1.51-CAPS and 1.85-HRM) were developed from pepper EST sequences corresponding to orthologs of tomato genes located at the 1,307,093–1,310,260, 1,501,436–1,506,288, and 1,846,027–1,849,951 bp positions of tomato chromosome 6, respectively. Application of and G05G1-HRM and 1.85-HRM to a subset (244 individuals) of a Chungyang F_2 population showed that the *Rf* gene was located within the pepper genomic region corresponding to the 1.71–1.85 Mb region on tomato chromosome 6 (Fig. 1b; Fig. S2). A total of 23 tomato genes were located in this region. Three further markers, G16-CAPS, G20-HRM, and G23-HRM, were developed based on pepper sequences corresponding to tomato genes located at the 1,795,940–1,797,216, 1,808,475–1,809,248, and 1,835,086–1,845,250 bp positions of tomato chromosome 6, respectively. Application of these markers to recombinants revealed that the *Rf* locus is located between G20-HRM and G23-HRM, which corresponds to the 1.81–1.84 Mb region in tomato (Fig. 1b; Fig. S2). Two tomato genes were found in this region, and only the gene encoding a homolog of a male sterility 5 (*Ms5*)-family protein in Arabidopsis could be aligned with the pepper genome sequence. However, two markers (ABHD1.5-SCAR, 4940-CAPS) flanking the pepper ortholog of tomato *Ms5* resulted in recombinants when they were applied to 1068 individuals of the Chungyang F_2 population (Fig. 1). These results imply that there might be no ortholog for the pepper *Rf* gene in the tomato genome and that the *Rf*-containing DNA region has uniquely evolved in pepper.

Anchoring developed markers on pepper genome sequences and integrating mapping information

To determine whether the pepper reference genomic DNA sequences from *C. annuum* CM334 (Kim et al. 2014) can be used in fine mapping to clone *Rf* candidate gene or not, we investigated *Rf* genotype in CM334. All the five F_1 individuals originated from test crosses with two CMS lines, Bukang A and Milyang A, respectively (see "Materials and methods"), were male-fertile showing that the CM334

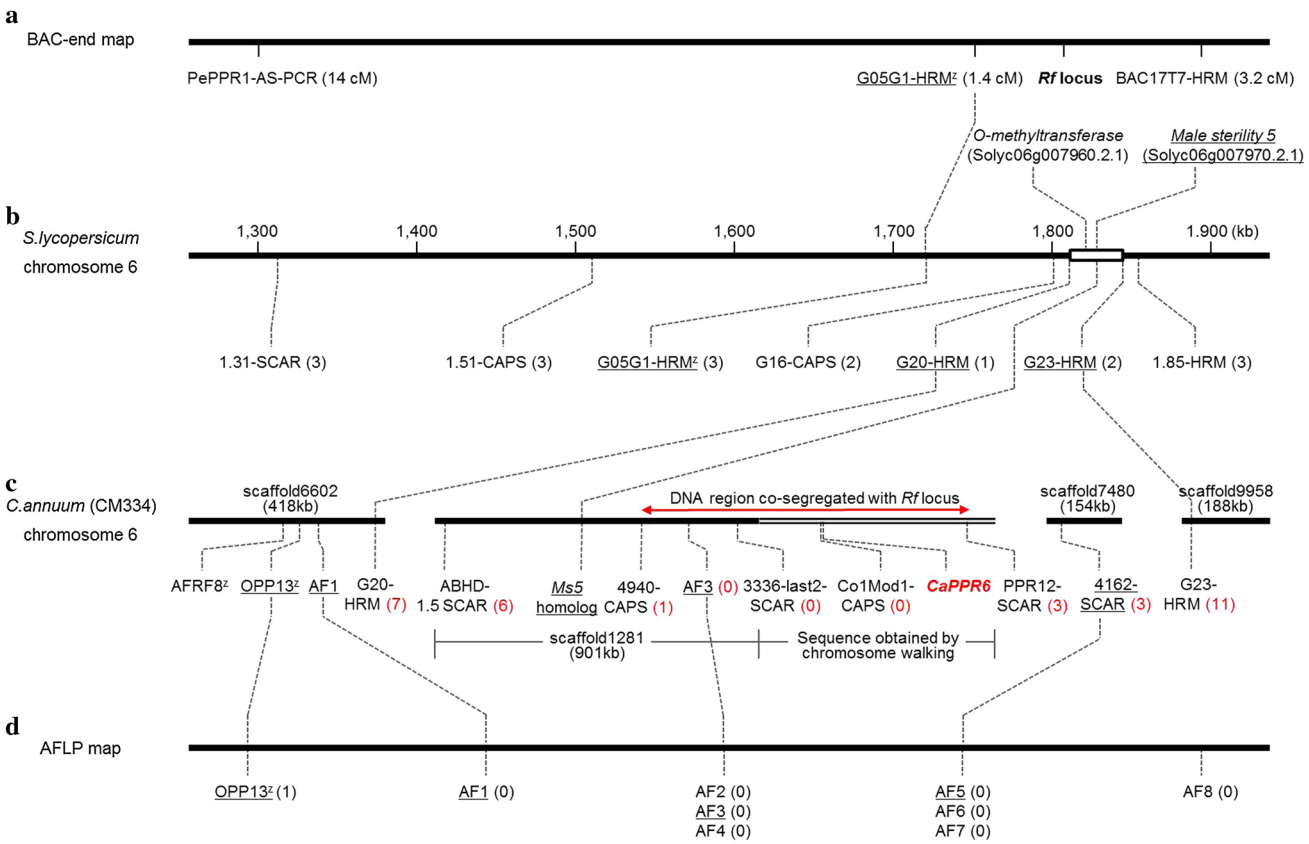
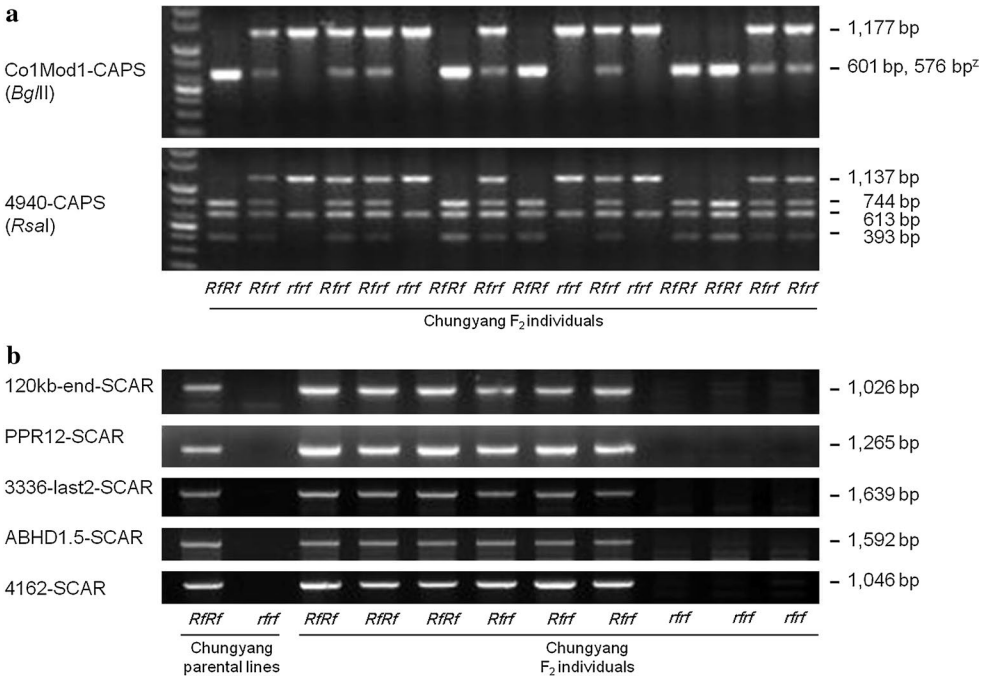


Fig. 1 Integration of mapping information and comparative analysis between genetic and physical maps. Markers reported in previous research are marked with *superscript letter z* (Kim 2005; Kim et al. 2006; Lee et al. 2008b). Markers or genes that were used to link different maps are underlined. **a** A genetic map from previous research (Jo et al. 2010) consisting of markers developed from end sequences of BAC clones screened with *RFL* probes. **b** Localization of pepper markers on a corresponding sequence of *S. lycopersicum* chromosome 6. Numbers of recombinants in a subset of the Chungyang F_2 mapping population (244 individuals) are in the parenthesis beside the marker names. The DNA region genetically corresponding to pepper *Rf* locus is depicted as an *open rectangle*. **c** Localization of markers on pepper scaffolds, and a sequence obtained during chromosome walking. Numbers of recombinants in the complete Chungyang F_2 mapping population (1068 individuals) are in parenthesis beside marker names and are *highlighted in red type*. **d** A genetic map consisting of markers developed by AFLP analysis (Fig. S1). Numbers of recombinants in 'TCMS' X 'Milyang K' F_2 mapping population (160 individuals) are in parenthesis beside marker names (color figure online)

contains the *Rf* gene. Markers developed by AFLP and comparative mapping were anchored to scaffold sequences of CM334. Four scaffold sequences containing the marker sequences were arrayed around the *Rf* locus, based on genetic distance between the anchored markers. Using the scaffold sequences, three new markers (4940-CAPS, 3336-last2-SCAR, 4162-SCAR) were developed and mapped using 1,068 individuals of the Chungyang F_2 population (Figs. 1c, 2). Scaffold 1281 was located on one side of the *Rf* locus, and the 4940-CAPS marker was found to be anchored to a 640-kb region of this scaffold. One recombinant was identified between 4940-CAPS markers and *Rf* locus. The 3336-last2-SCAR marker, which was developed from the right margin of the same scaffold, co-segregated with the *Rf* locus. Scaffold 7480 was located on the other side of the *Rf* locus, with the 4162-SCAR marker to one end of this scaffold. Three recombinants were identified between 4162-SCAR maker and *Rf* locus (Fig. 1c). Therefore, the genomic region containing the *Rf* locus was delimited to be between 4940-CAPS and 4162-SCAR (Figs. 1c, 2, 3). The locations of the markers on the genetic maps and the genomic DNA sequences showed clearly that the newly developed markers were closer to *Rf* than previously developed *Rf*-linked markers (Fig. 1; Kim 2005; Kim et al. 2006; Lee et al. 2008b). Comparative analysis between the genetic map of pepper and genomic sequence of tomato also showed that the *Rf*-containing region was unique to pepper (Fig. 1b, c).

Chromosome walking to define the DNA region co-segregating with *Rf*

Because of a high copy number of repetitive sequences, the genome sequence assembly at the *Rf*-containing region



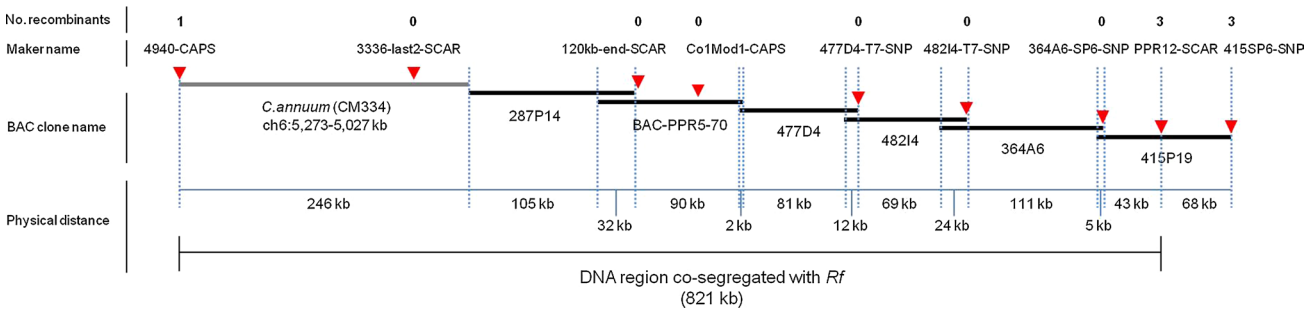


Fig. 2 Performance of markers developed from the co-segregating region, or its flanking region. **a** Performance of co-dominant markers. **b** Performance of dominant markers. Genotypes of individual samples are indicated at the bottom. [z] Two kinds of DNA fragments, which are 576 and 601 bp in length, are mixed in a single DNA band because of the small size difference between fragments

Fig. 3 Schematic diagram of chromosome walking to define the DNA region co-segregating with pepper *Rf*. The markers developed during chromosome walking are depicted as *red triangles* (color figure online)

might be unreliable. Therefore, six rounds of chromosome walking were carried out using a pepper BAC library (Yoo et al. 2003). The first chromosome walking cycle started from one end of scaffold 1281, which contains the marker co-segregating with *Rf* (3336-last2-SCAR, Fig. 3). Four SNP-based markers and two SCAR markers were developed to include the end sequences of each BAC clone from CM334 scaffold sequences. In the case that a developed marker in a selected BAC was found to co-segregate with *Rf*, the next BAC library screen was performed using the most distant end sequence of that BAC clone. The BAC clone '415P19' was isolated from the sixth round of chromosome walking, and from this, two markers (PPR12-SCAR, 415SP6-SNP) showing three recombinants from the *Rf* locus of the Chungyang F_2 population were identified (Fig. 3). Therefore, the region co-segregating with *Rf* was delimited to be between 4940-CAPS and PPR12-SCAR (developed from the BAC clone '415P19') (Figs. 1c, 3). Among those BAC clones identified as spanning the *Rf* locus, the 'BAC-PPR5-70' clone was identical to that selected in a previous study (Jo et al. 2010; Fig. 3).

Analysis of the DNA sequence co-segregating with *Rf* and candidate gene screening

The *Rf* locus-containing DNA sequence in the region between 4940-CAPS and PPR12-SCAR was 821 kb in length and contained numerous repeated sequences, mainly transposable elements or transposable element-like sequences, covering 82.3 % of the co-segregating region (Fig. 4a, b). By mapping reads (similarity >0.98) generated by sequencing the transcriptome of Bukang C anther tissue

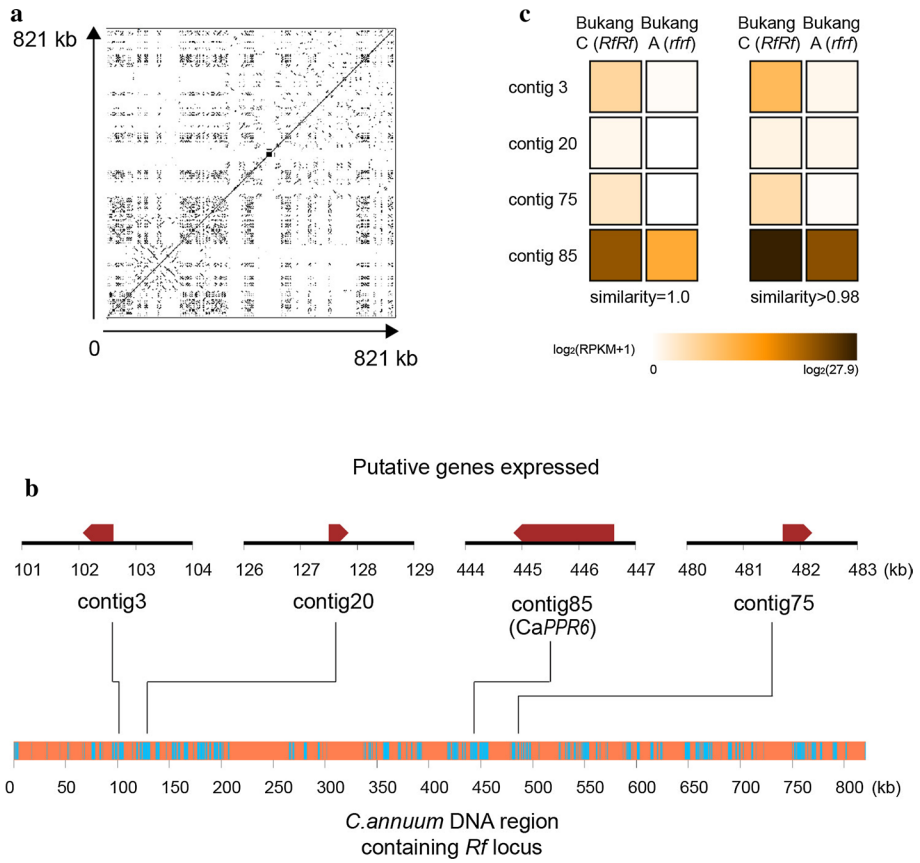


Fig. 4 Sequence analysis of the *Rf*-containing genomic region. **a** Dot plot analysis of repeat sequences in the *Rf* co-segregating region, visualized by the PIP maker program. **b** Heat map for RPKM of reads mapped on four contigs, localized on the co-segregating region in Bukang C and Bukang A. Read mapping was performed with two levels of stringency (similarity = 1.0 and similarity > 0.98). **c** Distribution of transposable elements and candidate genes in the co-segregating region. DNA regions covered by transposable elements are shown in *orange*, and the remaining region in *blue* (color figure online)

Table 1 Characteristics of transcripts from the pepper genomic region tightly linked with *Rf* locus

Contig name	Location on the co-segregating region		Contig length	Length of the longest *orf* in contig	Information on protein sequence in GenBank showing the highest similarity with ORF		
					GenBank ID	Characteristics	Length of CDS
Contig 3	101,892	102,670	779	534	XP_009631312.1	NAC domain-containing protein	699
Contig 20	127,366	127,875	510	342	XP_009601179.1	Uncharacterized	525
Contig 22	164,603	166,584	1837	273	XP_006352584.1	Putative rRNA methylase	858
Contig 24	166,675	166,990	316	51	–	–	
Contig 2	177,521	177,916	396	222	XP_009766553.1	NAC domain-containing protein	546
Contig 39	356,246	357,846	1077	261	–	–	
Contig 40	360,670	361,145	476	174	XP_009759827.1	Uncharacterized	1920
Contig 83	443,184	444,605	588	249	XP_009803299.1	Mitochondrial import receptor subunit TOM40-1-like	1170
Contig 85	444,723	446,627	588	589	NP_001306789.1	PPR protein	1776
Contig 75	481,504	482,547	805	513	XP_006362984.1	TIR1-like protein	1533
Contig 30	490,809	491,258	451	165	XP_009587921.1	Uncharacterized	1158

against the fine-mapped region, eleven transcript contigs with no similarity to transposable elements were identified (Table 1). Seven of the contigs (contigs 2, 22, 24, 30, 39, 40, and 83), and their corresponding sequences on CM334 BAC clones, contained no ORFs encoding peptides longer than 100 amino acids or covering more than 60 % of the

contig sequence in length. Although BLASTX analysis showed that five of these contigs encoded peptides similar to protein sequences in GenBank, stop codons were found in the middle of matching sequences, implying that they are not likely to encode intact or functional proteins. The remaining four contigs (contigs 3, 20, 75, and 85) encoded ORFs longer than 100 amino acids (Table 1; Fig. 4b). The *ORF*s in contigs 3 and 75 were identical to previously predicted coding sequences (CDS) in the CM334 genome project, while the *ORF* in contig 20 could not be matched with any CDS in the CM334 database (Kim et al. 2014). Contig 85 showed similarity with a CDS, but many mismatches were found (63 out of 1770 bp). Contigs 3, 75, and 85 encoded a NAC (NAM, ATAF1/2, and CUC2) domain protein, a TIR1 (transport inhibition response 1)-like protein, and a PPR protein, respectively, while contig 20 encoded an uncharacterized protein (Table 1). When the transcriptome sequence reads of Bukang C and Bukang A (CMS line) were mapped onto these four contigs under stringent conditions (similarity = 1.00, similarity > 0.98), many more sequences were mapped from Bukang C than from Bukang A (Fig. 4c). In addition, several sections of *CaPPR6* could not be mapped at all by transcriptome sequence reads of Bukang A (data not shown). These results indicate that the expression levels or sequences of these four genes were different between Bukang C and Bukang A. The difference in the number of reads mapped to contig 85 was greatly reduced in the milder mapping condition (similarity > 0.98) compared to the most stringent condition (similarity = 1.00). This finding implies that there is expression of sequences that are highly similar to contig 85 even in Bukang A. When expression levels of these genes were compared based on number of reads mapped, contig 85 had the highest expression of the four contigs (Fig. 4c).

Expression analysis and characterization of *CaPPR6*

The gene contained in contig 85 encoded a P-type PPR protein and was named *CaPPR6*. To investigate the expression of *CaPPR6* in plant tissues, RT-PCR was performed using a primer set specifically designed to amplify a portion of *CaPPR6* in CM334. Amplification products were found in four different plant tissues of Chungyang F_2 individuals with the *RfRf* genotype, but not in the same four tissues in the *rfrf* genotype (Fig. 5).

Based on transcriptome analysis, we suspected that multiple, highly similar homologs of *CaPPR6* would be present in both restorer and CMS lines. Using primers designed from the 5′ and 3′ end of *CaPPR6*, PCR was used to amplify two sequences including *CaPPR6* from the Chungyang restorer line and two homologs from the Chungyang CMS line (Fig. 6). Comparing protein sequences

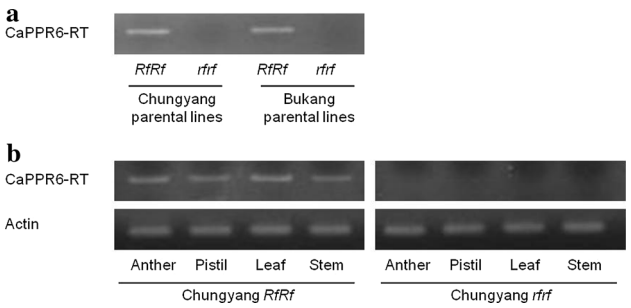


Fig. 5 PCR and RT-PCR outputs using a primer set that generates amplicons in Chungyang *C* and Bukang C, but not in Chungyang A and Bukang A. **a** PCR from genomic DNA from each parental line. **b** RT-PCR in different tissues of Chungyang F_2 individuals with *RfRf* and *rfrf* genotypes. *Actin* is included as an expression control

translated from homologous genes, CaPPR6 showed 93 % identity (96 % similarity) with CaPPR2 from the restorer line, while it was less similar (88 % identity/94 % similarity) to both CaPPR3 and CaPPR4 from the CMS line. CaPPR3 and CaPPR4 shared high similarity (98 % identity/98 % similarity).

Another pepper PPR protein sequence, encoded by a gene that was previously found to be located 14 cM from *Rf* (*CaPPR1*, Jo et al. 2010), was clearly distinguished from other PPR sequences in terms of protein length and sequence. Alignment of five PPR protein sequences revealed seven amino acids specific for CaPPR6 (Fig. 6). All five sequences contained 14 PPR motifs, with the level of conservation between them being lowest on the first, third and sixth amino acids across the 35 residues constituting each PPR motif (except for repeat 4, which consisted of 36 amino acids). Comparing the different PPR motif repeats, amino acid sequences in the tenth, eleventh, and twelfth repeats showed the lowest level of conservation (Fig. S3). When read mapping was carried out using *CaPPR6* as reference sequence (data not shown), it was not possible to map the DNA region corresponding to these repeats with high-throughput sequencing reads from Bukang A (CMS line).

Application of *Rf*-linked markers to breeding lines

Along with previously described markers, the newly developed *Rf*-linked markers from this research were applied to breeding lines provided by seed companies (Tables 2, 3). Application of the markers to 53 lines provided by Monsanto Korea revealed that CRF-SCAR, which had previously been tested on these lines, and the new markers developed from the region co-segregating with *Rf* showed a high ratio of successful genotyping (Table 2). Co1Mod1-CAPS, which was developed from the region closest to the

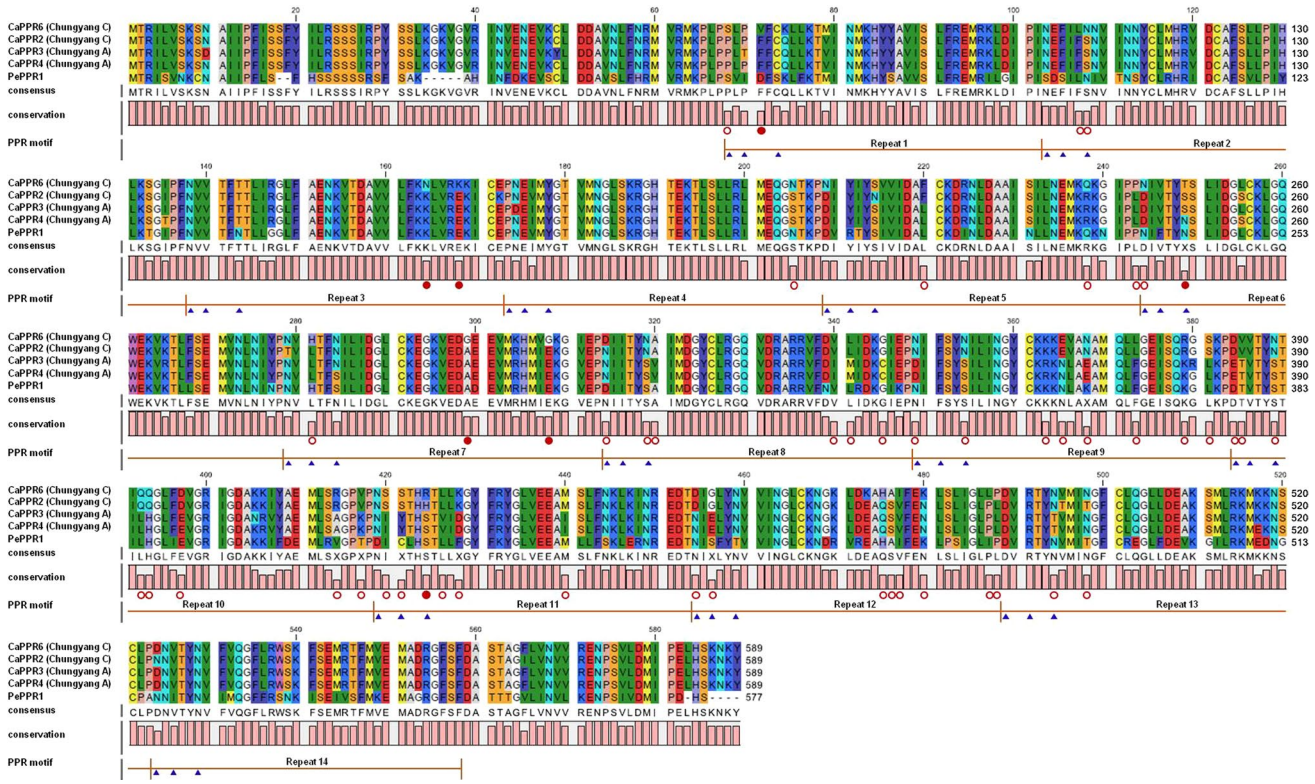


Fig. 6 Alignment of protein sequences encoded by *CaPPR6*, three homologs of *CaPPR6* isolated from Chungyang C (*CaPPR2*) and Chungyang A (*CaPPR3*, *CaPPR4*), and PePPR1 (Jo et al. 2010). The predicted PPR motifs are indicated and numbered below the sequence. *Open red circles* indicate residues that were identical in CaPPR6 and only one of homologs. *Filled red circles* indicate residues specific for CaPPR6. *Dark blue triangles* indicate the first, second, and third amino acids in each PPR motif (color figure online)

CaPPR6 gene (11 kb away), showed the highest accuracy (92.2 %), although several other newly developed markers including 4940-CAPS (88.5 %) and 3336-last2-SCAR (88.2 %) exhibited comparable accuracies. The 4940-CAPS marker was developed as a co-dominant marker in the Chungyang population, but it was not possible to generate amplicons in several lines, mainly those with the *rfrf* genotype. Other previously developed markers, OPP13-CAPS and BAC13T7-SCAR (Kim 2005; Lee et al. 2008b; Jo et al. 2010), and new markers developed from outside of the co-segregating region showed a very low ratio of successful genotyping (between 24.5 and 52.9 %; Table 2).

Application of the markers to 50 lines provided by Enza Zaden showed that the markers developed from the region tightly linked to *Rf* had a significantly higher ratio of successful genotyping (88.0 % for Co1Mod1-CAPS, 92.0 % for 3336-last2-SCAR) than CRF-SCAR (70.0 %) and 4940-CAPS (60.0 %). The accuracies for 3336-last2-SCAR and Co1Mod1-CAPS were equivalent considering that co-dominant genotyping was carried out for Co1Mod1-CAPS, whereas it was done dominantly for 3336-last2-SCAR (Table 3).

Discussion

Strategy for cloning PPR-type *Rf* genes using characteristics of *RFL* genes

To fine map the *Rf* locus in pepper, we used information obtained from AFLP and comparative mapping in combination with results from a candidate gene approach attempted in our previous study (Jo et al. 2010). Based on BAC clone and pepper genome sequences, the genomic region co-segregating with the *Rf* locus in our population was delimited to a sequence of 821 kb (Kim et al. 2014). This region contains a strong candidate gene for *Rf*, the PPR gene *CaPPR6*. *CaPPR6* is located on a BAC clone (BAC-PPR-5-70), which had previously been identified using a pepper EST sequence with high similarity to the petunia *Rf* gene, and typical characteristics of *RFL* genes (Jo et al. 2010). However, because of highly repetitive sequences and an incomplete whole genome sequence at that time, this BAC clone was neither grouped nor mapped by Jo et al. (2010). A comparative mapping approach using information about the tomato genome was not successful

Table 2 Application of seven markers linked to *Rf* to a panel of *C. annuum* breeding lines from Monsanto Korea

Markers	Marker haplotypes	Number of lines classified as homozygous for *Rf* (total 22 lines)	Number of lines classified as homozygous for *rf* (total 31 lines)	Ratio of successful genotyping (%)
OPP13-CAPS	OPP-haplotype 1	7[a]	0	24.5
	OPP-haplotype 2	1	6	
	OPP-haplotype 3	14	24	
	Heterozygous for OPP-haplotype 1 and 2	0	1	
CRF-SCAR	*RfRf* or *Rfrf*	20	4	88.7
	rfrf	2	27	
BAC13T7-SCAR	*RfRf* or *Rfrf*	5	9	50.0
	rfrf	16	20	
4940-CAPS[b]	*RfRf or Rfrf*	18	2	88.5
	Rfrf	1	1	
	rfrf	1	16	
	No amplification	1	12	
3336-last2-SCAR	*RfRf* or *Rfrf*	19	4	88.2
	rfrf	2	26	
Co1Mod1-CAPS	*RfRf*	20	1	92.2
	Rfrf	1	1	
	rfrf	1	27	
4162-SCAR	*RfRf* or *Rfrf*	19	22	52.9
	rfrf	2	8	
PPR12-SCAR	*RfRf* or *Rfrf*	19	28	41.2
	rfrf	2	2	

[a] Underlined numbers indicate the breeding lines whose phenotype was predicted correctly by each marker

[b] Although 4940-CAPS was developed as a co-dominant marker in the Chungyang population, it was not able to generate amplicons in many breeding lines, mostly those with the *rfrf* genotype. Therefore, when calculating the ratio of successful genotyping, the absence of amplicon in 4940-CAPS was counted as *rfrf*

in identifying the *Rf* gene because of a lack of synteny in the *Rf* region. Our current work demonstrates that hybridization-based BAC screening using *RFL* gene probes can be a straightforward and rapid strategy to isolate functional *Rf* genes, if linkage between one of the *RFL* genes and the *Rf* locus is confirmed by mapping analysis. *RFL* genes share high similarity between plant species, whereas they are clearly distinguished in their sequence from other *PPR* genes, even within a species (Dahan and Mireau 2013; Fujii et al. 2011). This feature enables specific BAC clones containing *RFL* genes to be isolated, and *RFL* genes may be screened based on known *Rf* gene sequences from other species. The utility of high-throughput short-read sequencing technologies to assemble PPR-type *Rf* genes is limited due to existence of many paralogs that show very high similarity in nucleotide sequence (Fig. 6). Indeed, the *CaPPR6* sequence was not found in the CM334 whole-genome sequence assembly, which was based on next-generation sequencing (Kim et al. 2014). The use of a BAC library can help to separate and classify each *RFL* gene. In addition, identified BAC clones can provide additional sequence information that can be used to develop markers to test for linkage with the restoration phenotype. The direct development of markers from *RFL* gene sequences is often hampered by the existence of highly similar sequences.

Isolation of a PPR-type *Rf* candidate gene

We identified four *Rf* candidate genes positioned on a genomic region tightly linked to the *Rf* phenotypes and specifically expressed in the restorer line. Except contig 85, all other contig are not likely to be *Rf*, since the *Rf* genes cloned thus far encode mitochondrion-targeted proteins that interact with CMS genes or are involved in metabolism or signaling in mitochondria (Chen and Liu 2014; Dahan and Mireau 2013). Contig 85 showed the highest level of expression among candidate genes and encoded a P-subfamily PPR protein (named CaPPR6 in this study). BLAST analysis for this protein against Arabidopsis protein database revealed the highest similarity to mitochondrion-targeted proteins that are classified as P-subfamily PPR proteins and encoded by genes clustered around a 4183–4326

Table 3 Application of five *Rf* markers to a panel of pepper breeding lines from Enza Zaden

Markers	Marker haplotypes	Number of lines classified as *RfRf* (total 24 lines)	Number of lines classified as *Rfrf* (total 12 lines)	Number of lines classified as *rfrf* (total 14 lines)	Ratio of successful genotyping (%)
OPP13-CAPS	OPP-haplotype 1	8[a]	0	0	24.0
	OPP-haplotype 2	1	0	0	
	OPP-haplotype 3	12	9	14	
	Heterozygous for haplotype 1 and 2	3	3	0	
CRF-SCAR	*RfRf* or *Rfrf*	15	8	2	70.0
	rfrf	9	4	12	
4940-CAPS[b]	*RfRf*	14	5	1	60.0
	Rfrf	1	4	1	
	rfrf	9	3	4	
	No amplification	0	0	8	
3336 last2	*RfRf* or *Rfrf*	21	12	1	92.0
SCAR	*rfrf*	3	0	13	
Co1Mod1-CAPS	*RfRf*	19	0	1	88.0
	Rfrf	2	12	0	
	rfrf	3	0	13	

[a] Underlined numbers indicate the breeding lines whose phenotype was predicted correctly by each marker

[b] Although 4940 CAPS was developed as co-dominant marker in the Chungyang population, it was not able to generate amplicons in many breeding lines, mostly those with the *rfrf* genotype. Therefore, when calculating the ratio of successful genotyping, the absence of amplicon in 4940-CAPS was counted as *rfrf*

or 23,176–23,509 kp region on Arabidopsis chromosome 1. Classified as *RFL* genes, the Arabidopsis P-subfamily PPR genes are densely clustered in this region and have been reported to be involved in processing mitochondrial transcripts (Fujii et al. 2011; Geddy and Brown 2007; Lurin et al. 2004; Jonietz et al. 2011; Holzle et al. 2011). CaPPR6 also shares high similarity with PPR-type Rf proteins isolated in other crops (63 % identity/76 % similarity to petunia Rf-PPR592, 33 % identity/54 % similarity to radish Rfo). The majority of *Rf* genes cloned so far belong to the P-subfamily of PPR genes, which show high similarity to *RFL* genes. Taking into account the predicted function of the encoded protein, the predicted organellar protein localization, the intact state of the gene, and the gene expression level in anthers, *CaPPR6* can be regarded as the strongest candidate for *Rf* among genes in the Rf-locus containing region of pepper genome.

Characteristics of *CaPPR6* and its homologs

The genomic region tightly linked to *Rf* was not syntenic with the tomato genome and consisted of many transposable elements and truncated genes (Fig. 6a, c; Table 2). Kim et al. (2014) demonstrated that retrotransposon proliferation, especially of members of the *Gypsy* family, resulted in extreme expansion of the pepper genome compared to tomato. Transposable element coverage in the *Rf*-co-segregating region (82.3 %) exceeded that in the CM334 whole genome (76.4 %). In this region, retrotransposons in the *Gypsy* family occupied the highest proportion among repeated sequences, indicating that extensive retrotransposon insertions may have contributed to the pepper-specific evolution of this region (data not shown). Outside of the defined co-segregation region, three additional PPR pseudogenes, which were truncated or contained numerous stop codons, were found on the BAC clone 415P19, suggesting that the *Rf* gene resides within a genomic region containing multiple PPR genes.

The candidate gene *CaPPR6* remained intact and was normally transcribed, clearly distinguishing it from other truncated genes. In addition, this gene was located far from the nearest PPR genes (PPR pseudogenes on BAC clone 415P19) unlike *RFL*s in rice, Arabidopsis and petunia, which are densely clustered (Lurin et al. 2004; Bentolila et al. 2002; Wang et al. 2006). It has been suggested that *RFL*s undergo active evolution and are maintained or lost according an 'arms race' with mitochondrial target genes following birth-and-death processes similar to disease resistance genes (Dahan and Mireau 2013; Touzet and Budar 2004). Therefore, the protection of *CaPPR6* from the severe modifications or rearrangements that are predicted to occur on surrounding sequences implies that the role of *CaPPR6* has remained undiminished during evolution.

The manner in which *RFL* genes evolve makes it difficult to elucidate the recessive allele of the PPR gene. However, we isolated two genes homologous to *CaPPR6* in a CMS line, as well as one *RFL* gene showing high similarity to *CaPPR6* in a restorer line. Alignment of homologous RFL protein sequences, such as petunia Rf protein (Bentolila et al. 2002), showed that all of them contained 14 PPR repeats and divergence among the proteins was highest for the first, third, and sixth amino acids in each PPR repeat. Fujii et al. (2011) reported that *RFL* genes have undergone diversifying selection, particularly on these three amino acids. Although CaPPR6 and CaPPR2 were very similar (identity: 93 %), there were 10 polymorphic amino acids between them on the first or sixth residue of the PPR motifs (Fig. 6). It has been suggested that the first and sixth amino acids of PPR motifs directly interact with target RNA sequences (Barkan et al. 2012; Barkan and Small 2014). Taking this into account, the results presented here imply that divergence of these amino acids may have caused the differentiation of mitochondrial target genes between CaPPR6 and its homologs. In silico analysis of PPR sequences is needed to predict interactions between CaPPR6 (and its homologs) and transcripts of CMS candidate genes, including *orf507* and *Ψatp6-2*. *In vitro* experiments are also required confirm these interactions (Kim and Kim 2006; Kim et al. 2007; Li et al. 2013) and to suggest a mechanism for fertility restoration.

Development of markers with broader applicability to molecular *Rf* breeding

For efficient molecular breeding of CGMS in pepper, development of markers that are tightly linked to *Rf* and broadly applicable for a wide range of breeding progenitors has been attempted. In addition, as partial restoration and instability of restoration have been observed, particularly in sweet peppers (Kim et al. 2013; Shifriss 1997), the markers were also applied in discrimination between lines with a stable restorer from those with unstable restorers (Lee et al. 2008a, b; Min et al. 2008, 2009). For example, OPP13 and the other markers derived from OPP13 have been developed (Kim 2005; Kim et al. 2013; Lee et al. 2008b; Min et al. 2008, 2009; Zhang et al. 2000). Although the markers were shown to be 0.37 (Zhang et al. 2000) to 1.1 (Lee et al. 2008b) cM from *Rf* depending on *Rf*-segregation populations, application of these markers, for example OPP13-CAPS, to a broad range of breeding lines showed very low accuracy (below 30 %). CRF-SCAR showed the highest accuracy (70.0–89.1 %) among the *Rf*-linked markers (Jo et al. 2010 and in this study). The markers developed from the co-segregating sequence region defined in this study perform significantly better than any of the previously developed markers in terms of their applicability to a broad range of breeding lines (Tables 2, 3). They also showed high accuracy in a mapping population (Table S1; Table S2; Fig. 1c). Of particular note, Co1Mod1-CAPS, which is located near *CaPPR6*, was developed as co-dominant marker and is expected to be useful in practical molecular breeding. Although the applicability of markers tends to correlate with their proximity to *CaPPR6*, those developed from the co-segregating region were not able to correctly detect the genotype in more than one-tenth of breeding lines. It is possible that there are multiple *Rf* genes. In rice, two *Rf* genes were identified in three CMS cytoplasms including WA-CMS (*Rf3*, *Rf4*; Luo et al. 2013), BT-CMS (*Rf1a*, *Rf1b*; Wang et al. 2006), and HL-CMS (*Rf5*, *Rf6*; Hu et al. 2012). In addition, different haplotypes can exist even for an allele of the *Rf* gene. Ohgami et al. (2016) reported that there are at least three haplotypes for the recessive allele of *Rf* (*rf1*) in sugar beet. Further studies on the classification of haplotypes for *CaPPR6* and genetic analysis of *Rf* for lines in which markers were not applicable is needed to investigate these possibilities.

Author contribution statement YDJ participated in design of the study, marker development, genetic mapping, chromosome walking, and bioinformatics analysis. HY pooled the BAC library and carried out RT-PCR. LJH performed comparative mapping. MP and DC screened and provided pepper genome sequences. ACB and PJD supported AFLP analysis. HIC supported analysis of repeat sequence and candidate gene screening. SG and BK edited the manuscript. BCK participated in conception of the study, discussion, and revision of the manuscript. All authors read and approved the final version of the manuscript.

Acknowledgments This research was supported by the Golden Seed Project (213002-04-3-CG900), the Ministry of Agriculture, Food and Rural Affairs (MAFRA), the Ministry of Oceans and Fisheries (MOF), the Rural Development Administration (RDA), and the Korea Forest Service (KFS), Republic of Korea, and a grant (710001-07) from the Vegetable Breeding Research Center through the Agriculture, Food and Rural Affairs Research Center Support Program, Ministry of Agriculture, Food and Rural Affairs.

Compliance with ethical standards

Conflict of interest This research was supported by the Golden Seed Project (213002-04-3-CG900), the Ministry of Agriculture, Food and Rural Affairs (MAFRA), the Ministry of Oceans and Fisheries (MOF), the Rural Development Administration (RDA), and the Korea Forest Service (KFS), Republic of Korea, and a grant (710001-07) from the Vegetable Breeding Research Center through the Agriculture, Food and Rural Affairs Research Center Support Program, Ministry of Agriculture, Food and Rural Affairs.

References

Barkan A, Small I (2014) Pentatricopeptide repeat proteins in plants. Annu Rev Plant Biol 65:415–442

Barkan A, Rojas M, Fujii S, Yap A, Chong YS, Bond CS, Small I (2012) A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. PLoS Genet 8:e1002910

Barr CM, Fishman L (2010) Cytoplasmic male sterility in *Mimulus* hybrids has pleiotropic effects on corolla and pistil traits. Heredity (Edinb) 106:886–893

Bentolila S, Alfonso AA, Hanson MR (2002) A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. Proc Natl Acad Sci USA 99:10887–10892

Brown GG, Formanova N, Jin H, Wargachuk R, Dendy C, Patil P, Laforest M, Zhang J, Cheung WY, Landry BS (2003) The radish *Rfo* restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. Plant J 35:262–272

Chen L, Liu YG (2014) Male sterility and fertility restoration in crops. Annu Rev Plant Biol 65:579–606

Cui X, Wise RP, Schnable PS (1996) The *rf2* nuclear restorer gene of male-sterile T-cytoplasm maize. Science 272:1334–1336

Dahan J, Mireau H (2013) The *Rf* and *Rf*-like PPR in higher plants, a fast-evolving subclass of PPR genes. RNA Biol 10:1469–1476

Desloire S, Gherbi H, Laloui W, Marhadour S, Clouet V, Cattolico L, Falentin C, Giancola S, Renard M, Budar F, Small I, Caboche M, Delourme R, Bendahmane A (2003) Identification of the fertility restoration locus, *Rfo*, in radish, as a member of the pentatricopeptide-repeat protein family. EMBO Rep 4:588–594

Fujii S, Toriyama K (2009) Suppressed expression of Retrograde-Regulated Male Sterility restores pollen fertility in cytoplasmic male sterile rice plants. Proc Natl Acad Sci USA 106:9513–9518

Fujii S, Bond CS, Small ID (2011) Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. Proc Natl Acad Sci USA 108:1723–1728

Geddy R, Brown GG (2007) Genes encoding pentatricopeptide repeat (PPR) proteins are not conserved in location in plant genomes and may be subject to diversifying selection. BMC Genom 8:130

Gulyas G, Pakozdi K, Lee JS, Hirata Y (2006) Analysis of fertility restoration by using cytoplasmic male-sterile red pepper (*Capsicum annuum* L.) lines. Breed Sci 56:331–334

Hanson MR, Bentolila S (2004) Interactions of mitochondrial and nuclear genes that affect male gametophyte development. Plant Cell 16(Suppl):S154–169

Holzle A, Jonietz C, Torjek O, Altmann T, Binder S, Forner J (2011) A RESTORER OF FERTILITY-like PPR gene is required for 5′-end processing of the *nad4* mRNA in mitochondria of *Arabidopsis thaliana*. Plant J 65:737–744

Hu J, Wang K, Huang W, Liu G, Gao Y, Wang J, Huang Q, Ji Y, Qin X, Wan L, Zhu R, Li S, Yang D, Zhu Y (2012) The rice pentatricopeptide repeat protein RF5 restores fertility in Hong-Lian cytoplasmic male-sterile lines via a complex with the glycine-rich protein GRP162. Plant Cell 24:109–122

Huang W, Yu C, Hu J, Wang L, Dan Z, Zhou W, He C, Zeng Y, Yao G, Qi J, Zhang Z, Zhu R, Chen X, Zhu Y (2015) Pentatricopeptide-repeat family protein RF6 functions with hexokinase 6 to rescue rice cytoplasmic male sterility. Proc Natl Acad Sci USA 112:14984–14989

Itabashi E, Iwata N, Fujii S, Kazama T, Toriyama K (2011) The fertility restorer gene, *Rf2*, for lead rice-type cytoplasmic male sterility of rice encodes a mitochondrial glycine-rich protein. Plant J 65:359–367

Jeong HJ, Jo YD, Kang BC (2010) Identification of *Capsicum* species using SNP markers based on high resolution melting analysis. Genome 53:1029–1040

Jo YD, Kim YM, Park MN, Yoo JH, Park M, Kim BD, Kang BC (2010) Development and evaluation of broadly applicable markers for *Restorer-of-fertility* in pepper. Mol Breed 25:187–201

Jonietz C, Forner J, Hildebrandt T, Binder S (2011) RNA PROCESSING FACTOR3 is crucial for the accumulation of mature *ccmC* transcripts in mitochondria of *Arabidopsis* accession Columbia. Plant Physiol 157:1430–1439

Kazama T, Toriyama K (2003) A pentatricopeptide repeat-containing gene that promotes the processing of aberrant *atp6* RNA of cytoplasmic male-sterile rice. FEBS Lett 544:99–102

Kim DS (2005) Development of RAPD and AFLP markers linked to *fertility restorer* (*Rf*) gene in chili pepper (*Capsicum annuum* L.). Thesis, Seoul National University

Kim DH, Kim BD (2006) The organization of mitochondrial *atp6* gene region in male fertile and CMS lines of pepper (*Capsicum annuum* L.). Curr Genet 49:59–67

Kim DS, Kim DH, Yoo JH, Kim BD (2006) Cleaved amplified polymorphic sequence and amplified fragment length polymorphism markers linked to the fertility restorer gene in chili pepper (*Capsicum annuum* L.). Mol Cells 21:135–140

Kim DH, Kang JG, Kim BD (2007) Isolation and characterization of the cytoplasmic male sterility-associated *orf456* gene of chili pepper (*Capsicum annuum* L.). Plant Mol Biol 63:519–532

Kim HJ, Baek KH, Lee SW, Kim J, Lee BW, Cho HS, Kim WT, Choi D, Hur CG (2008) Pepper EST database: comprehensive in silico tool for analyzing the chili pepper (*Capsicum annuum*) transcriptome. BMC Plant Biol 8:101

Kim YM, Jo YD, Kwon JK, Kim IT, Kang BC (2013) Characterization and inheritance of a novel thermo-sensitive restoration of cytoplasmic male sterility in *Capsicum annuum*. Sci Hortic 164:512–520

Kim S, Park M, Yeom SI, Kim YM, Lee JM, Lee HA, Seo E, Choi J, Cheong K, Kim KT, Jung K, Lee GW, Oh SK, Bae C, Kim SB, Lee HY, Kim SY, Kim MS, Kang BC, Jo YD, Yang HB, Jeong HJ, Kang WH, Kwon JK, Shin C, Lim JY, Park JH, Huh JH, Kim JS, Kim BD, Cohen O, Paran I, Suh MC, Lee SB, Kim YK, Shin Y, Noh SJ, Park J, Seo YS, Kwon SY, Kim HA, Park JM, Kim HJ, Choi SB, Bosland PW, Reeves G, Jo SH, Lee BW, Cho HT, Choi HS, Lee MS, Yu Y, Do Choi Y, Park BS, van Deynze A, Ashrafi H, Hill T, Kim WT, Pai HS, Ahn HK, Yeam I, Giovannoni JJ, Rose JK, Sorensen I, Lee SJ, Kim RW, Choi IY, Choi BS, Lim JS, Lee YH, Choi D (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. Nat Genet 46:270–278

Klein RR, Klein PE, Mullet JE, Minx P, Rooney WL, Schertz KF (2005) Fertility restorer locus Rf1 [corrected] of sorghum (*Sorghum bicolor* L.) encodes a pentatricopeptide repeat protein not present in the colinear region of rice chromosome 12. Theor Appl Genet 111:994–1012

Koizuka N, Imai R, Fujimoto H, Hayakawa T, Kimura Y, Kohno-Murase J, Sakai T, Kawasaki S, Imamura J (2003) Genetic characterization of a pentatricopeptide repeat protein gene, *orf687*, that restores fertility in the cytoplasmic male-sterile Kosena radish. Plant J 34:407–415

Komori T, Ohta S, Murai N, Takakura Y, Kuraya Y, Suzuki S, Hiei Y, Imaseki H, Nitta N (2004) Map-based cloning of a fertility restorer gene, *Rf-1*, in rice (*Oryza sativa* L.). Plant J 37:315–325

Kubo T, Kitazaki K, Matsunaga M, Kagami H, Mikami T (2011) Male sterility-inducing mitochondrial genomes: how do they differ? Crit Rev Plant Sci 30:378–400

Lee J, Yoon JB, Park HG (2008a) A CAPS marker associated with the partial restoration of cytoplasmic male sterility in chili pepper (*Capsicum annuum* L.). Mol Breed 21:95–104

Lee J, Yoon JB, Park HG (2008b) Linkage analysis between the *partial restoration* (*pr*) and the *restorer-of-fertility* (*Rf*) loci in pepper cytoplasmic male sterility. Theor Appl Genet 117:383–389

Li J, Pandeya D, Jo YD, Liu WY, Kang BC (2013) Reduced activity of ATP synthase in mitochondria causes cytoplasmic male sterility in chili pepper. Planta 237:1097–1109

Liu C, Ma N, Wang PY, Fu N, Shen HL (2013) Transcriptome sequencing and *de novo* analysis of a cytoplasmic male sterile line and its near-isogenic restorer line in chili pepper (*Capsicum annuum* L.). PLoS ONE 8:e65209

Luo D, Xu H, Liu Z, Guo J, Li H, Chen L, Fang C, Zhang Q, Bai M, Yao N, Wu H, Wu H, Ji C, Zheng H, Chen Y, Ye S, Li X, Zhao X, Li R, Liu YG (2013) A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in rice. Nat Genet 45:573–577

Lurin C, Andres C, Aubourg S, Bellaoui M, Bitton F, Bruyere C, Caboche M, Debast C, Gualberto J, Hoffmann B, Lecharny A, Le Ret M, Martin-Magniette ML, Mireau H, Peeters N, Renou JP, Szurek B, Taconnat L, Small I (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell 16:2089–2103

Matsuhira H, Kagami H, Kurata M, Kitazaki K, Matsunaga M, Hamaguchi Y, Hagihara E, Ueda M, Harada M, Muramatsu A, Yui-Kurino R, Taguchi K, Tamagake H, Mikami T, Kubo T (2012) Unusual and typical features of a novel *restorer-of-fertility* gene of sugar beet (*Beta vulgaris* L.). Genetics 192:1347–1358

Min WK, Lim H, Lee YP, Sung SK, Kim BD, Kim S (2008) Identification of a third haplotype of the sequence linked to the *Restorer-of-fertility* (*Rf*) gene and its implications for male-sterility phenotypes in peppers (*Capsicum annuum* L.). Mol Cells 25:20–29

Min WK, Kim S, Sung SK, Kim BD, Lee S (2009) Allelic discrimination of the *Restorer-of-fertility* gene and its inheritance in peppers (*Capsicum annuum* L.). Theor Appl Genet 119:1289–1299

Ohgami T, UchiyamaD Ue S, Yui-Kurino R, Yoshida Y, Kamei Y, Kuroda Y, Taguchi K, Kubo T (2016) Identification of molecular variants of the nonrestoring restorer-of-fertility 1 allele in sugar beet (*Beta vulgaris* L.). Theor Appl Genet 129:675–688

Peterson PA (1958) Cytoplasmically inherited male sterility in *Capsicum*. Am Nat 92:111–119

Qin C, Yu C, Shen Y, Fang X, Chen L, Min J, Cheng J, Zhao S, Xu M, Luo Y, Yang Y, Wu Z, Mao L, Wu H, Ling-Hu C, Zhou H, Lin H, Gonzalez-Morales S, Trejo-Saavedra DL, Tian H, Tang X, Zhao M, Huang Z, Zhou A, Yao X, Cui J, Li W, Chen Z, Feng Y, Niu Y, Bi S, Yang X, Li W, Cai H, Luo X, Montes-Hernandez S, Leyva-Gonzalez MA, Xiong Z, He X, Bai L, Tan S, Tang X, Liu D, Liu J, Zhang S, Chen M, Zhang L, Zhang L, Zhang Y, Liao W, Zhang Y, Wang M, Lv X, Wen B, Liu H, Luan H, Zhang Y, Yang S, Wang X, Xu J, Li X, Li S, Wang J, Palloix A, Bosland PW, Li Y, Krogh A, Rivera-Bustamante RF, Herrera-Estrella L, Yin Y, Yu J, Hu K, Zhang Z (2014) Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. Proc Natl Acad Sci USA 111:5135–5140

Shifriss C (1997) Male sterility in pepper (*Capsicum annuum* L.). Euphytica 93:83–88

Small ID, Peeters N (2000) The PPR motif—a TPR-related motif prevalent in plant organellar proteins. Trends Biochem Sci 25:46–47

Tomato Genome C (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635–641

Touzet P, Budar F (2004) Unveiling the molecular arms race between two conflicting genomes in cytoplasmic male sterility? Trends Plant Sci 9:568–570

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M et al (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Wang Z, Zou Y, Li X, Zhang Q, Chen L, Wu H, Su D, Chen Y, Guo J, Luo D, Long Y, Zhong Y, Liu YG (2006) Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. Plant Cell 18:676–687

Xu XB, Liu ZX, Zhang DF, Liu Y, Song WB, Li JS, Dai JR (2009) Isolation and analysis of rice *Rf1*-orthologus PPR genes co-segregating with *Rf3* in maize. Plant Mol Biol Rep 27:511–517

Yoo EY, Kim S, Kim YH, Lee CJ, Kim BD (2003) Construction of a deep coverage BAC library from *Capsicum annuum*, 'CM334'. Theor Appl Genet 107:540–543

Zhang BX, Huang SW, Yang GM, Guo JZ (2000) Two RAPD markers linked to a major fertility restorer gene in pepper. Euphytica 113:155–161